

## ABSTRACT

From the last decade, there has been an explosion of web and social media content (*SMC*) in the virtual world. It poses many challenges starting from processing of vast amounts of data, to the presence of colloquial/code-mixed language to excessive prevalence of fake and exaggerated news etc. While the list of problems is long, *SMC* also provides many opportunities. For example, understanding sentiment toward products and services etc. In this thesis, we intend to evaluate the capability of social signals gathered from the online crowd to build various applications to solve social and NLP challenges posed by *SMC*. As a first step toward our objective, with the help of social signals gathered from the online crowd, we characterize the spread of exaggeration and classify the users based on their propensity to share exaggerated news article. As a subsequent step, we show how gathered social signals can be used to solve an NLP task in a monolingual setting, i.e., exaggeration detection, and in a multi-lingual context, i.e., likelihood of word borrowing. We achieve state-of-the-art performance in both of these tasks.